PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68	A2	(11) International Publication Number: **WO 99/00516**
		(43) International Publication Date: 7 January 1999 (07.01.99)

(21) International Application Number: PCT/GB98/01862

(22) International Filing Date: 25 June 1998 (25.06.98)

(30) Priority Data:
9713469.6 25 June 1997 (25.06.97) GB

(71) Applicant *(for all designated States except US)*: MEDICAL RESEARCH COUNCIL [GB/GB]; 20 Park Crescent, London W1N 4AL (GB).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: LOVELL-BADGE, Robin [GB/GB]; National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA (GB). PEVNY, Laryssa, H. [GB/GB]; National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA (GB). SMITH, Austin [GB/GB]; Centre for Genome Research, University of Edinburgh, East Mains Road, Edinburgh EH9 3WQ (GB).

(74) Agents: MASCHIO, Antonio et al.; D. Young & Co., 21 New Fetter Lane, London EC4A 1DA (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD. SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: NEURONAL STEM CELL GENE

(57) Abstract

The invention relates to a method for isolating a neuroblastic cell from a population of cells comprising the steps of: (a) detecting the expression of the *Sox1* gene in the cells; and (b) sorting the cells to isolate those cells expressint the *Sox1* gene, as well as to a method for causing a cell to become committed to the neural pathway comprising administering SOX1 thereto.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

NEURONAL STEM CELL GENE

The present invention relates to a method of marking, selecting and generating neuronal stem sells from tissues. In particular, the invention relates to the use of the *Sox1* gene for the generation of neuroblasts.

SOX proteins constitute a family of transcription factors related to the mammalian testis determining factor SRY through homology within their HMG box DNA binding domains. In DNA binding studies, SOX proteins exhibit sequence specific binding; however, unlike most transcription factors, binding occurs in the minor groove resulting in the induction of a dramatic bend within the DNA helix. SOX proteins can induce transcription of reporter constructs in vitro and display properties of both classical transcription factors and architectural components of chromatin (reviewed by Peveny and Lovell-Badge (1997) Curr. Opin. Genetics and Development, 7:338-344).

Members of the *Sox* gene family are expressed in a variety of embryonic and adult tissues, where they appear to be responsible for the development and/or elaboration of particular cell lineages. *Sry* is transiently expressed in the precursor Sertoli cells of the XY genital ridge and is responsible for triggering development of the male phenotype (reviewed by Lovell-Badge and Hacker, (1995) Phil. Trans. R. Soc. Lond. B 350:205-214). Thus, the lack of *Sry* results in XY females and XX males. *Sox9* is expressed in immature chondrocytes and male gonads; mutations in the human SOX9 gene are associated with Campomelic Dysplasia, a human skeletal malformation syndrome, and XY female sex reversal. *Sox4 is* expressed in many tissues and a null mutation of the gene in mouse results in the absence of mature B cells and heart malformations. *Xsox17* genes are involved in endoderm formation in Xenopus embryos. These functional analyses Suggest that *Sox* genes function in cell fate decisions in diverse developmental pathways.

A subfamily of *Sox* genes, that includes *Sox1, Sox2 and Sox3*, shows expression profiles during vertebrate embryogenesis that suggest the genes could function in the control of' cell fate decisions within the early developing nervous system. *Sox2 and Sox3* begin to be expressed at preimplantation and epiblast stages respectively, and are then restricted to the neuroepithelium. *Sox1* appears only at around the stage of neural induction.

The molecular mechanisms controlling neural induction and determination have begun to be elucidated. The identification by cellular and biochemical methods, of secreted molecules involved in neural induction illustrates the important role of the environment in specifying cell identity. In addition, a number of transcription factors have been isolated which play important roles in the specification and differentiation of neural cell lineages. For example, the characterisation of vertebrate homologues of Drosophila proneural and neurogenic genes, which control neural specification in the fly, has revealed analogous molecular mechanisms in vertebrate neural cell fate determination and differentiation. In an Drosophila, the expression of basic helix-loop-helix transcription factors of the AS-C complex confirms neural potential on groups of ectodermal cells. Miss expression of a transcription factors involved in a neural cell fate determination is observed to cause abnormalities in neural development.

It is known that *Sox1* expression appears only at around the stage of neural induction in the embryo. The role of SOX1 in embryogenesis is, however, not known.

Summary of the Invention

It is shown herein that *Sox1* expression correlates with the formation of the neural plate. Moreover, the onset of *Sox1* expression in embryonal carcinoma cells is shown to be dependent on neural induction. Upregulation of *Sox1* expression is itself sufficient to impart a neural fate on pluripotent cells.

In a first aspect of the present invention, there is provided a method for isolating a neuroblastic cell from a population of cells comprising the steps of:

(a) detecting the expression of the *Sox1* gene in the cells; and

(b) sorting the cells to isolate those cells expressing the *Sox1* gene.

As set forth in the following description, the *Sox1* gene, which encodes SOX1, is responsible for the specification of neuroblast or neuronal stem cells, as well as acting as a marker for such cells. Expression of *Sox1* is responsible for the generation of the neuroblastic cell type, which *in vivo* is capable of differentiating into the many different cells and ganglia of the CNS. Moreover, *Sox1* is a unique marker for neuroblasts.

Cells which are identified as expressing this gene, for example by binding to anti-SOX1 antibodies, by activation of SOX1 dependent ligand-receptor systems or by detection with antisense nucleic acids specific for *Sox1* mRNA, are pluripotent neuroblasts. Such cells can be identified in early embryonic tissue or adult CNS material. Cells can be sorted by affinity techniques, or by cell sorting (such as fluorescence-activated cell sorting) where they are labelled with a suitable label, such as a fluorophore conjugated to or part of, for example, an antisense nucleic acid molecule or an immunoglobulin.

According to a second aspect of the invention, neuroblast cells can be actively sorted from other cell types by detecting the expression of SOX1 *in vivo* using a reporter system. Thus, for example, the invention provides a method for isolating a neuroblastic cell from a population of cells, comprising the steps of:

(a) transfecting the population of cells with a genetic construct comprising a coding sequence encoding a detectable marker operatively linked to the *Sox1* control regions;

(b) detecting the cells which express the selectable marker; and

(c) sorting the cells which express the selectable marker from the population of cells.

As before, the selectable marker may be any selectable entity, but is preferably a fluorescent or luminescent marker which may be detected and sorted by automated cell sorting approaches. For example, the marker may be GFP or luciferase. Other useful markers include those which are expressed in the cell membrane, thus facilitating cell sorting by affinity means.

The genetic construct according to the invention may comprise any promoter and enhancer elements as required, so long as the overall control remains sensitive to SOX1; in other words, no expression of the marker coding sequence should take place in the absence of SOX1. The regulatory sequences of the SOX1 gene are known in the art and have been described in the literature cited herein and incorporated herein by reference; at least, however, the construct of the invention will comprise a SOX1 binding site. Preferably, the SOX1 control elements are used in their entirety; however, other promoter and enhancer elements may be substituted where they remain under the influence of SOX1.

The selectable marker will only be expressed in neuroblastic cells because only these cells express SOX1, which is required for transcription from the *Sox1* control sequences. Preferably, therefore, the expression system used to express the selectable marker is not leaky and expresses a minimal amount of the marker in the absence of SOX1. Techniques for transforming cells with coding genetic constructs according to the invention, detecting the marker and sorting cells accordingly are known in the art.

The present invention, in a third aspect, provides the use of the *Sox1* coding sequence to transform precursor cells and thereby differentiate neuroblast cells therefrom. Accordingly, there is provided a method for differentiating one or more

neuroblastic cells from one or mores pluripotent precursor cell, comprising the steps of:

(a) transforming the pluripotent precursor cell(s) with a genetic construct comprising a *Sox1* coding sequence operatively linked to suitable control sequences; and

(b) culturing the cell(s) so as to allow expression of the *Sox1* coding sequence, thereby inducing the cell to differentiate into a neuroblast.

Suitable control sequences for use in the third aspect of the invention are known in the art and may include inducible or constitutive control sequences. Inducible control sequences have the advantage that *Sox1* expression may be switched off when desired, for example once the cell is to be differentiated into another neural cell. Moreover, once the expression of exogenous *Sox1* has been switched off, successfully differentiated neuroblasts may be identified by virtue of the continued expression of the endogenous *Sox1* gene.

Precursor cells may be, for example, ES cells, such as human ES cells and cells with similar pluripotent properties derived from germ cells (EG cells). More specific neuronal pluripotent precursors or direct neuroblast precursors may also be employed.

Neuroblasts obtained according to the invention may be employed in a number of ways. Of course, the expression of *Sox1* has important implications for the study of neural differentiation; the generation and selection of neuroblasts will provide material for basic research.

Moreover, the invention has medical and diagnostic applications. The detection of *Sox1* expressing cells is important in clinical neurology and in diagnosing and treating cancers of the nervous system. Accordingly, the invention provides a method for detecting the presence of a neuroblast as described above for diagnostic purposes.

Neural stem cells are also useful for the treatment of neurological disorders, especially for repair of accidentally induced trauma in the CNS or for the correction of congenital or pathological diseases of the CNS.

Moreover, in applications involving somatic gene therapy designed to correct a genetic defect in nervous tissue, the removal, treatment and replacement of pluripotent neuroblasts which are actively dividing has clear advantages, providing a constant source of modified neural cells to permanently treat the targeted defect. *Sox1* control sequences may be used specifically to direct transgene expression in neuroblast cells where this is desired. Moreover, gene expression can be directed to neural cell types differentiated from neuroblasts by the use of other control sequences, such as NF-1 control sequences which direct expression of NF-1 in mature neurons *in vivo*.

A significant advantage of the methods described herein is that a patient in need of treatment for a neurological disorder can act as a self-donor. In other words, cells may be isolated from the patient and either sorted to extract neuroblasts, or treated in order to differentiate neuroblasts as described, from specific or general precursors.

Detailed Description of the Invention

The present invention relates to a method for isolating, or producing cells which are committed to the neural fate. Accordingly, the term neuroblast, as used herein, refers to any cell or cell line which has commenced differentiation along the neural pathways.

The isolation of neuroblastic cells from populations of cells is desirable, in order to obtain cells which are committed to neural pathways, but are not terminally differentiated. Such cells are useful in the study of neuronal differentiation, and in

the treatment of diseases such as neurodegenerative diseases, and neural damage, for example occasioned by trauma. Thus, typical populations of cells from which neuroblastic cells may be differentiated include cell populations derived from the CNS of mammals, such as humans, including CNS from adult and foetal sources. Moreover, cell populations derived from tissue cultures may be employed for the isolation of neuroblastic cells.

It has been determined that SOX1 expression is closely associated with the acquisition of neural fate by the ectoderm, both in vitro and in vivo. *In vitro* SOX1 expression is initiated within 24 hours after the addition of retinoic acid to pluripotent EC cell aggregates coincident with the induction of neuroepithelial markers such as NESTIN, *Mash1* and *Wnt1*. In mouse and rat embryos expression is restricted to cells of the antero/distal ectoderm. Previous fate mapping studies indicated that this region of the epiblast constitutes the promordium of the nervous system.

Expression of SOX1 is detected throughout the cells of the neural plate and early neural tube along its entire anteroposterior axis. The early and uniform expression of SOX1 throughout the presumptive CNS indicates that SOX1 is activated by neural inducing signals and lends support to the proposal of a two step response of the ectoderm to organiser signals in generating a nervous system: neutralisation followed by regionalisation.

Expression of this *Sox* gene subfamily has been evolutionarily conserved. The Drosophila (Nambu and Nambu 1996; Russel et al., 1996) zebrafish (Vriz et al., 1996) and avian (Unwanogho et al., 1995; Streit et al., 1997; Rex et al., 1997) putative orthologues of *Sox1*, *Sox2* and *Sox3* all show expression throughout the neural primordium. Thus, this subfamily of Sox genes represents a novel group of transcription factors which can serve as general early neuroepithelial markers.

In order to isolate neuroblastic cells, the present invention provides for the detection of *Sox1* therein. As used herein, *Sox1* may be derived from any source, including mammalian sources, avian sources and other vertebrate sources. *Sox1* may also be derived from invertebrate sources.

Sox1 has been cloned from human, chicken and mouse. The sequence of chicken, mouse and human *Sox1* is set forth in SEQ ID.s numbers 1 to 3 herein.

The preferred sequence encoding *Sox1* is that encoding human *Sox1* and having substantially the same nucleotide sequence as the sequence in SEQ ID No. 3, with the nucleic acid having the same sequence as the sequence in SEQ ID No. 3 being most preferred. As used herein, nucleotide sequences which are substantially the same share at least about 90% identity. However, in the case of splice variants having e.g. an additional exon sequence homology may be lower.

The nucleic acids of the invention, whether used as probes or otherwise, are preferably substantially homologous to the sequence of human *Sox1* as shown in SEQ ID No. 3. As used herein, "homology" means that the two entities share sufficient characteristics for the skilled person to determine that they are similar in origin and function. Preferably, homology is used to refer to sequence identity. Thus, *Sox1* sequences according to the invention preferably retain substantial sequence identity human *Sox1*.

"Substantial homology", where homology indicates sequence identity, means more than 40% sequence identity, preferably more than 45% sequence identity and most preferably a sequence identity of 50% or more, as judged by direct sequence alignment and comparison.

Sequence homology (or identity) may be determined using any suitable homology algorithm, using for example default parameters. Advantageously, the BLAST algorithm is employed, with parameters set to default values. The BLAST

algorithm is described in detail at http://www.ncbi.nih.gov/BLAST/blast_help.html, which is incorporated herein by reference. The search parameters are defined as follows, and are advantageously set to the defined default parameters.

Advantageously, "substantial homology" when assessed by BLAST equates to sequences which match with an EXPECT value of at least 7, preferablyat least 9 and most preferably 10 or more. The default threshold for EXPECT in BLAST searching is ususally 10.

BLAST (Basic Local Alignment Search Tool) is the heuristic search algorithm employed by the programs blastp, blastn, blastx, tblastn, and tblastx; these programs ascribe significance to their findings using the statistical methods of Karlin and Altschul (see http://www.ncbi.nih.gov/BLAST/blast_help.html) with a few enhancements. The BLAST programs were tailored for sequence similarity searching, for example to identify homologues to a query sequence. The programs are not generally useful for motif-style searching. For a discussion of basic issues in similarity searching of sequence databases, see Altschul et al. (1994) Nature Genetics 6:119-129.

The five BLAST programs available at http://www.ncbi.nlm.nih.gov perform the following tasks:

blastp compares an amino acid query sequence against a protein sequence database;

blastn compares a nucleotide query sequence against a nucleotide sequence database;

blastx compares the six-frame conceptual translation products of a nucleotide query sequence (both strands) against a protein sequence database;

tblastn compares a protein query sequence against a nucleotide sequence database dynamically translated in all six reading frames (both strands).

tblastx compares the six-frame translations of a nucleotide query sequence against the six-frame translations of a nucleotide sequence database.

BLAST uses the following search parameters:

HISTOGRAM Display a histogram of scores for each search; default is yes. (See parameter H in the BLAST Manual).

DESCRIPTIONS Restricts the number of short descriptions of matching sequences reported to the number specified; default limit is 100 descriptions. (See parameter V in the manual page). See also EXPECT and CUTOFF.

ALIGNMENTS Restricts database sequences to the number specified for which high-scoring segment pairs (HSPs) are reported; the default limit is 50. If more database sequences than this happen to satisfy the statistical significance threshold for reporting (see EXPECT and CUTOFF below), only the matches ascribed the greatest statistical significance are reported. (See parameter B in the BLAST Manual).

EXPECT The statistical significance threshold for reporting matches against database sequences; the default value is 10, such that 10 matches are expected to be found merely by chance, according to the stochastic model of Karlin and Altschul (1990). If the statistical significance ascribed to a match is greater than the EXPECT threshold, the match will not be reported. Lower EXPECT thresholds are more stringent, leading to fewer chance matches being reported. Fractional values are acceptable. (See parameter E in the BLAST Manual).

CUTOFF Cutoff score for reporting high-scoring segment pairs. The default value is calculated from the EXPECT value (see above). HSPs are reported for a database sequence only if the statistical significance ascribed to them is at least as high as would be ascribed to a lone HSP having a score equal to the CUTOFF value. Higher CUTOFF values are more stringent, leading to fewer chance matches being reported. (See parameter S in the BLAST Manual). Typically, significance thresholds can be more intuitively managed using EXPECT.

MATRIX Specify an alternate scoring matrix for BLASTP, BLASTX, TBLASTN and TBLASTX. The default matrix is BLOSUM62 (Henikoff & Henikoff, 1992). The valid alternative choices include: PAM40, PAM120, PAM250 and IDENTITY. No alternate scoring matrices are available for BLASTN; specifying the MATRIX directive in BLASTN requests returns an error response.

STRAND Restrict a TBLASTN search to just the top or bottom strand of the database sequences; or restrict a BLASTN, BLASTX or TBLASTX search to just reading frames on the top or bottom strand of the query sequence.

FILTER Mask off segments of the query sequence that have low compositional complexity, as determined by the SEG program of Wootton & Federhen (1993) Computers and Chemistry 17:149-163, or segments consisting of short-periodicity internal repeats, as determined by the XNU program of Claverie & States (1993) Computers and Chemistry 17:191-201, or, for BLASTN, by the DUST program of Tatusov and Lipman (see http://www.ncbi.nlm.nih.gov). Filtering can eliminate statistically significant but biologically uninteresting reports from the blast output (e.g., hits against common acidic-, basic- or proline-rich regions), leaving the more biologically interesting regions of the query sequence available for specific matching against database sequences.

Low complexity sequence found by a filter program is substituted using the letter "N" in nucleotide sequence (e.g., "NNNNNNNNNNNNN") and the letter "X" in

protein sequences (e.g., "XXXXXXXXX"). Users may turn off filtering by using the "Filter" option on the "Advanced options for the BLAST server" page.

Filtering is only applied to the query sequence (or its translation products), not to database sequences. Default filtering is DUST for BLASTN, SEG for other programs.

It is not unusual for nothing at all to be masked by SEG, XNU, or both, when applied to sequences in SWISS-PROT, so filtering should not be expected to always yield an effect. Furthermore, in some cases, sequences are masked in their entirety, indicating that the statistical significance of any matches reported against the unfiltered query sequence should be suspect.

NCBI-gi Causes NCBI gi identifiers to be shown in the output, in addition to the accession and/or locus name.

Most preferably, sequence comparisons are conducted using the simple BLAST search algorithm provided at http://www.ncbi.nlm.nih.gov/BLAST.

Preferably, the invention makes use of fragments of the *Sox1*-encoding sequence. Fragments of the nucleic acid sequence of a few nucleotides in length, preferably 5 to 150 nucleotides in length, are especially useful as probes.

Exemplary nucleic acids can alternatively be characterised as those nucleotide sequences which encode a *Sox1* protein and hybridise to the DNA sequences set forth SEQ ID No. 3, or a selected fragment of said DNA sequence. Preferred are such sequences encoding *Sox1* which hybridise under high-stringency conditions to the sequence of SEQ ID No. 3.

Stringency of hybridisation refers to conditions under which polynucleic acids hybrids are stable. Such conditions are evident to those of ordinary skill in the field.

As known to those of skill in the art, the stability of hybrids is reflected in the melting temperature (Tm) of the hybrid which decreases approximately 1 to 1.5°C with every 1% decrease in sequence homology. In general, the stability of a hybrid is a function of sodium ion concentration and temperature. Typically, the hybridisation reaction is performed under conditions of higher stringency, followed by washes of varying stringency.

As used herein, high stringency refers to conditions that permit hybridisation of only those nucleic acid sequences that form stable hybrids in 1 M Na+ at 65-68 °C. High stringency conditions can be provided, for example, by hybridisation in an aqueous solution containing 6x SSC, 5x Denhardt's, 1 % SDS (sodium dodecyl sulphate), 0.1 Na+ pyrophosphate and 0.1 mg/ml denatured salmon sperm DNA as non specific competitor. Following hybridisation, high stringency washing may be done in several steps, with a final wash (about 30 min) at the hybridisation temperature in 0.2 - 0.1x SSC, 0.1 % SDS.

Moderate stringency refers to conditions equivalent to hybridisation in the above described solution but at about 60-62°C. In that case the final wash is performed at the hybridisation temperature in 1x SSC, 0.1 % SDS.

Low stringency refers to conditions equivalent to hybridisation in the above described solution at about 50-52°C. In that case, the final wash is performed at the hybridisation temperature in 2x SSC, 0.1 % SDS.

It is understood that these conditions may be adapted and duplicated using a variety of buffers, e.g. formamide-based buffers, and temperatures. Denhardt's solution and SSC are well known to those of skill in the art as are other suitable hybridisation buffers (see, e.g. Sambrook, et al., eds. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York or Ausubel, et al., eds. (1990) Current Protocols in Molecular Biology, John Wiley & Sons,

Inc.). Optimal hybridisation conditions have to be determined empirically, as the length and the GC content of the hybridising pair also play a role.

Advantageously, the invention moreover provides nucleic acid sequence which are capable of hybridising, under stringent conditions, to a fragment of SEQ. ID. No. 3. Preferably, the fragment is between 15 and 50 bases in length. Advantageously, it is about 25 bases in length.

As will be appreciated by those skilled in the art, the redundancy of the genetic code allows the design of a large number of sequences encoding human *Sox1*. Any of these sequences may be useful for expressing SOX1 as described below. An advantage of the use of a sequence encoding human SOX1 which is not the human *Sox1* sequence is that the mRNA produced has a different sequence to that of the endogenous *Sox1* mRNA, and may thus be distinguished therefrom. Antisense oligonucleotides may be designed which are capable of selectively inhibiting the expression of either endogenous or exogenous *Sox1* genes. Degenerate sequences encoding human SOX1 are set forth in SEQ. ID. No. 5.

Given the guidance provided herein, nucleic acids encoding *Sox1* are obtainable according to methods well known in the art. For example, a nucleic acid encoding *Sox1* is obtainable by chemical synthesis, using polymerase chain reaction (PCR) or by screening a genomic library or a suitable cDNA library prepared from a source believed to possess *Sox1* and to express it at a detectable level.

Chemical methods for synthesis of a nucleic acid of interest are known in the art and include triester, phosphite, phosphoramidite and H-phosphonate methods, PCR and other autoprimer methods as well as oligonucleotide synthesis on solid supports. These methods may be used if the entire nucleic acid sequence of the nucleic acid is known, or the sequence of the nucleic acid complementary to the coding strand is available. Alternatively, if the target amino acid sequence is known, one may infer

potential nucleic acid sequences using known and preferred coding residues for each amino acid residue.

An alternative means to isolate a gene encoding *Sox1* is to use PCR technology as described e.g. in section 14 of Sambrook et al., 1989. This method requires the use of oligonucleotide probes that will hybridise to *Sox1* nucleic acid. Strategies for selection of oligonucleotides are described below.

Libraries are screened with probes or analytical tools designed to identify the gene of interest or the protein encoded by it. For cDNA expression libraries suitable means include monoclonal or polyclonal antibodies that recognise and specifically bind to *Sox1*; oligonucleotides of about 20 to 80 bases in length that encode known or suspected *Sox1* cDNA from the same or different species; and/or complementary or homologous cDNAs or fragments thereof that encode the same or a hybridising gene. Appropriate probes for screening genomic DNA libraries include, but are not limited to oligonucleotides, cDNAs or fragments thereof that encode the same or hybridising DNA; and/or homologous genomic DNAs or fragments thereof.

A nucleic acid encoding *Sox1* may be isolated by screening suitable cDNA or genomic libraries under suitable hybridisation conditions with a probe, i.e. a nucleic acid disclosed herein including oligonucleotides derivable from the sequences set forth in SEQ ID NO. 3. Suitable libraries are commercially available or can be prepared e.g. from cell lines, tissue samples, and the like.

As used herein, a probe is e.g. a single-stranded DNA or RNA that has a sequence of nucleotides that includes between 10 and 50, preferably between 15 and 30 and most preferably at least about 20 contiguous bases that are the same as (or the complement of) an equivalent or greater number of contiguous bases set forth in SEQ ID No. 3. The nucleic acid sequences selected as probes should be of sufficient length and sufficiently unambiguous so that false positive results are minimised. The nucleotide sequences are usually based on conserved or highly

homologous nucleotide sequences or regions of *Sox1*. The nucleic acids used as probes may be degenerate at one or more positions. The use of degenerate oligonucleotides may be of particular importance where a library is screened from a species in which preferential codon usage in that species is not known.

Preferred regions from which to construct probes include 5' and/or 3' coding sequences, sequences predicted to encode ligand binding sites, and the like. For example, either the full-length cDNA clone disclosed herein or fragments thereof can be used as probes. Preferably, nucleic acid probes of the invention are labelled with suitable label means for ready detection upon hybridisation. For example, a suitable label means is a radiolabel. The preferred method of labelling a DNA fragment is by incorporating α^{32}P dATP with the Klenow fragment of DNA polymerase in a random priming reaction, as is well known in the art. Oligonucleotides are usually end-labelled with γ^{32}P-labelled ATP and polynucleotide kinase. However, other methods (e.g. non-radioactive) may also be used to label the fragment or oligonucleotide, including e.g. enzyme labelling, fluorescent labelling with suitable fluorophores and biotinylation.

After screening the library, e.g. with a portion of DNA including substantially the entire *Sox1*-encoding sequence or a suitable oligonucleotide based on a portion of said DNA, positive clones are identified by detecting a hybridisation signal; the identified clones are characterised by restriction enzyme mapping and/or DNA sequence analysis, and then examined, e.g. by comparison with the sequences set forth herein, to ascertain whether they include DNA encoding a complete *Sox1* (i.e., if they include translation initiation and termination codons). If the selected clones are incomplete, they may be used to rescreen the same or a different library to obtain overlapping clones. If the library is genomic, then the overlapping clones may include exons and introns. If the library is a cDNA library, then the overlapping clones will include an open reading frame. In both instances, complete clones may be identified by comparison with the DNAs and deduced amino acid sequences provided herein.

It is envisaged that Sox 1-encoding sequences can be readily modified by nucleotide substitution, nucleotide deletion, nucleotide insertion or inversion of a nucleotide stretch, and any combination thereof. Such mutants can be used e.g. to produce a SOX1 mutant that has an amino acid sequence differing from the SOX1 sequences as found in nature. Mutagenesis may be predetermined (site-specific) or random. A mutation which is not a silent mutation must not place sequences out of reading frames and preferably will not create complementary regions that could hybridise to produce secondary mRNA structure such as loops or hairpins.

Sorting of cells, based upon detection of expression of the *Sox1* gene, may be performed by any technique known in the art, as exemplified above. For example, the cells may be sorted by flow cytometry or FACS. For a general reference, see Flow Cytometry and Cell Sorting: A Laboratory Manual (1992) A. Radbruch (Ed.), Springer Laboratory, New York.

Flow cytometry is a powerful method for studying and purifying cells. It has found wide application, particularly in immunology and cell biology: however, the capabilities of the FACS can be applied in many other fields of biology. The acronym F.A.C.S. stands for Fluorescence Activated Cell Sorting, and is used interchangeably with "flow cytometry". The principle of FACS is that individual cells, held in a thin stream of fluid, are passed through one or more laser beams, causing light to be scattered and fluorescent dyes to emit light at various frequencies. Photomultiplier tubes (PMT) convert light to electrical signals, which are interpreted by software to generate data about the cells. Sub-populations of cells with defined characteristics can be identified and automatically sorted from the suspension at very high purity ($\sim 100\%$).

FACS machines collect fluorescence signals in one to several channels corresponding to different laser excitation and fluorescence emission wavelengths. Fluorescent labelling allows the investigation of many aspects of cell structure and

function. The most widely used application is immunofluorescence: the staining of cells with antibodies conjugated to fluorescent dyes such as fluorescein and phycoerythrin. This method is often used to label molecules on the cell surface, but antibodies can also be directed at targets within the cell. In direct immunofluorescence, an antibody to a particular molecule, the SOX1 polypeptide, is directly conjugated to a fluorescent dye. Cells can then be stained in one step. In indirect immunofluorescence, the primary antibody is not labelled, but a second fluorescently conjugated antibody is added which is specific for the first antibody: for example, if the anti-SOX1 antibody is a mouse IgG, then the second antibody could be a rat or rabbit antibody raised against mouse IgG.

FACS can be used to measure gene expression in cells transfected with recombinant DNA encoding SOX1. This can be achieved directly, by labelling of the protein product, or indirectly by using a reporter gene in the construct. Examples of reporter genes are β-galactosidase and Green Fluorescent Protein (GFP). β-galactosidase activity can be detected by FACS using fluorogenic substrates such as fluorescein digalactoside (FDG). FDG is introduced into cells by hypotonic shock, and is cleaved by the enzyme to generate a fluorescent product, which is trapped within the cell. One enzyme can therefore generate a large amount of fluorescent product. Cells expressing GFP constructs will fluoresce without the addition of a substrate. Mutants of GFP are available which have different excitation frequencies, but which emit fluorescence in the same channel. In a two-laser FACS machine, it is possible to distinguish cells which are excited by the different lasers and therefore assay two transfections at the same time.

Alternative means of cell sorting may also be employed. For example, the invention comprises the use of nucleic acid probes complementary to *Sox1* mRNA. Such probes can be used to identify cells expressing *Sox1* individually, such that they may subsequently be sorted either manually, or using FACS sorting. Nucleic acid probes complementary to *Sox1* mRNA may be prepared according to the

teaching set forth above, using the general procedures as described by Sambrook et al (1989).

In a preferred embodiment, the invention comprises the use of an antisense nucleic acid molecule, complementary to *Sox1* mRNA, conjugated to a fluorophore which may be used in FACS cell sorting.

Suitable imaging agents for use with FACS may be delivered to the cells by any suitable technique, including simple exposure thereto in cell culture, delivery of transiently expressing nucleic acids by viral or non- viral vector means, liposome-mediated transfer of nucleic acids or imaging agents, and the like.

The invention, in certain embodiments, includes antibodies specifically recognising and binding to SOX1. For example, such antibodies may be generated against the SOX1 having the amino acid sequences set forth in SEQ ID No. 4. Alternatively, SOX1 or SOX1 fragments (which may also be synthesised by in vitro methods) are fused (by recombinant expression or an in vitro peptidyl bond) to an immunogenic polypeptide and this fusion polypeptide, in turn, is used to raise antibodies against a SOX1 epitope.

Anti-SOX1 antibodies may be recovered from the serum of immunised animals. Monoclonal antibodies may be prepared from cells from immunised animals in the conventional manner.

The antibodies of the invention are useful for identifying SOX1 in neural cells expressing *Sox1*, in accordance with the present invention.

Antibodies according to the invention may be whole antibodies of natural classes, such as IgE and IgM antibodies, but are preferably IgG antibodies. Moreover, the invention includes antibody fragments, such as Fab, F(ab')2, Fv and ScFv. Small fragments, such Fv and ScFv, possess advantageous properties for diagnostic and

therapeutic applications on account of their small size and consequent superior tissue distribution.

The antibodies may comprise a label. Especially preferred are labels which allow the imaging of the antibody in neural cells in vivo. Such labels may be radioactive labels or radioopaque labels, such as metal particles, which are readily visualisable within tissues. Moreover, they may be fluorescent labels or other labels which are visualisable in tissues and which may be used for cell sorting.

Recombinant DNA technology may be used to improve the antibodies of the invention. Thus, chimeric antibodies may be constructed in order to decrease the immunogenicity thereof in diagnostic or therapeutic applications. Moreover, immunogenicity may be minimised by humanising the antibodies by CDR grafting [see European Patent Application 0 239 400 (Winter)] and, optionally, framework modification.

Antibodies according to the invention may be obtained from animal serum, or, in the case of monoclonal antibodies or fragments thereof, produced in cell culture. Recombinant DNA technology may be used to produce the antibodies according to established procedure, in bacterial or preferably mammalian cell culture. The selected cell culture system preferably secretes the antibody product.

Therefore, the present invention includes a process for the production of an antibody according to the invention comprising culturing a host, e.g. E. coli or a mammalian cell, which has been transformed with a hybrid vector comprising an expression cassette comprising a promoter operably linked to a first DNA sequence encoding a signal peptide linked in the proper reading frame to a second DNA sequence encoding said protein, and isolating said protein.

Multiplication of hybridoma cells or mammalian host cells in vitro is carried out in suitable culture media, which are the customary standard culture media, for example

Dulbecco's Modified Eagle Medium (DMEM) or RPMI 1640 medium, optionally replenished by a mammalian serum, e.g. foetal calf serum, or trace elements and growth sustaining supplements, e.g. feeder cells such as normal mouse peritoneal exudate cells, spleen cells, bone marrow macrophages, 2-aminoethanol, insulin, transferrin, low density lipoprotein, oleic acid, or the like. Multiplication of host cells which are bacterial cells or yeast cells is likewise carried out in suitable culture media known in the art, for example for bacteria in medium LB, NZCYM, NZYM, NZM, Terrific Broth, SOB, SOC, 2 x YT, or M9 Minimal Medium, and for yeast in medium YPD, YEPD, Minimal Medium, or Complete Minimal Dropout Medium.

In vitro production provides relatively pure antibody preparations and allows scale-up to give large amounts of the desired antibodies. Techniques for bacterial cell, yeast or mammalian cell cultivation are known in the art and include homogeneous suspension culture, e.g. in an airlift reactor or in a continuous stirrer reactor, or immobilised or entrapped cell culture, e.g. in hollow fibres, microcapsules, on agarose microbeads or ceramic cartridges.

Large quantities of the desired antibodies can also be obtained by multiplying mammalian cells in vivo. For this purpose, hybridoma cells producing the desired antibodies are injected into histocompatible mammals to cause growth of antibody-producing tumours. Optionally, the animals are primed with a hydrocarbon, especially mineral oils such as pristane (tetramethyl-pentadecane), prior to the injection. After one to three weeks, the antibodies are isolated from the body fluids of those mammals. For example, hybridoma cells obtained by fusion of suitable myeloma cells with antibody-producing spleen cells from Balb/c mice, or transfected cells derived from hybridoma cell line Sp2/0 that produce the desired antibodies are injected intraperitoneally into Balb/c mice optionally pre-treated with pristane, and, after one to two weeks, ascitic fluid is taken from the animals.

The cell culture supernatants are screened for the desired antibodies, preferentially by immunofluorescent staining of cells expressing SOX1, by immunoblotting, by an enzyme immunoassay, e.g. a sandwich assay or a dot-assay, or a radioimmunoassay.

For isolation of the antibodies, the immunoglobulins in the culture supernatants or in the ascitic fluid may be concentrated, e.g. by precipitation with ammonium sulphate, dialysis against hygroscopic material such as polyethylene glycol, filtration through selective membranes, or the like. If necessary and/or desired, the antibodies are purified by the customary chromatography methods, for example gel filtration, ion-exchange chromatography, chromatography over DEAE-cellulose and/or (immuno-)affinity chromatography, e.g. affinity chromatography with SOX1 protein or with Protein-A.

The invention further concerns hybridoma cells secreting the monoclonal antibodies of the invention. The preferred hybridoma cells of the invention are genetically stable, secrete monoclonal antibodies of the invention of the desired specificity and can be activated from deep-frozen cultures by thawing and recloning.

The invention also concerns a process for the preparation of a hybridoma cell line secreting monoclonal antibodies directed SOX1, characterised in that a suitable mammal, for example a Balb/c mouse, is immunised with purified SOX1 protein, an antigenic carrier containing purified SOX1 or with cells bearing SOX1, antibody-producing cells of the immunised mammal are fused with cells of a suitable myeloma cell line, the hybrid cells obtained in the fusion are cloned, and cell clones secreting the desired antibodies are selected. For example spleen cells of Balb/c mice immunised with cells bearing SOX1 are fused with cells of the myeloma cell line PAI or the myeloma cell line Sp2/0-Ag14, the obtained hybrid cells are screened for secretion of the desired antibodies, and positive hybridoma cells are cloned.

Preferred is a process for the preparation of a hybridoma cell line, characterised in that Balb/c mice are immunised by injecting subcutaneously and/or intraperitoneally between 10 and 107 and 108 cells of human tumour origin which express SOX1 containing a suitable adjuvant several times, e.g. four to six times, over several months, e.g. between two and four months, and spleen cells from the immunised mice are taken two to four days after the last injection and fused with cells of the myeloma cell line PAI in the presence of a fusion promoter, preferably polyethylene glycol. Preferably the myeloma cells are fused with a three- to twentyfold excess of spleen cells from the immunised mice in a solution containing about 30 % to about 50 % polyethylene glycol of a molecular weight around 4000. After the fusion the cells are expanded in suitable culture media as described hereinbefore, supplemented with a selection medium, for example HAT medium, at regular intervals in order to prevent normal myeloma cells from overgrowing the desired hybridoma cells.

The invention also concerns recombinant DNAs comprising an insert coding for a heavy chain variable domain and/or for a light chain variable domain of antibodies directed to the extracellular domain of SOX1 as described hereinbefore. By definition such DNAs comprise coding single stranded DNAs, double stranded DNAs consisting of said coding DNAs and of complementary DNAs thereto, or these complementary (single stranded) DNAs themselves.

Furthermore, DNA encoding a heavy chain variable domain and/or for a light chain variable domain of antibodies directed SOX1 can be enzymatically or chemically synthesised DNA having the authentic DNA sequence coding for a heavy chain variable domain and/or for the light chain variable domain, or a mutant thereof. A mutant of the authentic DNA is a DNA encoding a heavy chain variable domain and/or a light chain variable domain of the above-mentioned antibodies in which one or more amino acids are deleted or exchanged with one or more other amino acids. Preferably said modification(s) are outside the CDRs of the heavy chain variable domain and/or of the light chain variable domain of the antibody. Such a

mutant DNA is also intended to be a silent mutant wherein one or more nucleotides are replaced by other nucleotides with the new codons coding for the same amino acid(s). Such a mutant sequence is also a degenerated sequence. Degenerated sequences are degenerated within the meaning of the genetic code in that an unlimited number of nucleotides are replaced by other nucleotides without resulting in a change of the amino acid sequence originally encoded. Such degenerated sequences may be useful due to their different restriction sites and/or frequency of particular codons which are preferred by the specific host, particularly E. coli, to obtain an optimal expression of the heavy chain murine variable domain and/or a light chain murine variable domain.

The term mutant is intended to include a DNA mutant obtained by in vitro mutagenesis of the authentic DNA according to methods known in the art.

For the assembly of complete tetrameric immunoglobulin molecules and the expression of chimeric antibodies, the recombinant DNA inserts coding for heavy and light chain variable domains are fused with the corresponding DNAs coding for heavy and light chain constant domains, then transferred into appropriate host cells, for example after incorporation into hybrid vectors.

The invention therefore also concerns recombinant DNAs comprising an insert coding for a heavy chain murine variable domain of an antibody directed SOX1 fused to a human constant domain g, for example γ1, γ2, γ3 or γ4, preferably γ1 or γ4. Likewise the invention concerns recombinant DNAs comprising an insert coding for a light chain murine variable domain of an antibody directed to SOX1 fused to a human constant domain κ or λ, preferably κ.

In another embodiment the invention pertains to recombinant nucleic acids wherein the heavy chain variable domain and the light chain variable domain are linked by way of a DNA insert coding for a spacer group, optionally comprising a signal sequence facilitating the processing of the antibody in the host cell and/or a DNA

coding for a peptide facilitating the purification of the antibody and/or a DNA coding for a cleavage site and/or a DNA coding for a peptide spacer and/or a DNA coding for an effector molecule, such as a label.

According to a further aspect, and as referred to above, neuroblastic cells may be actively sorted from other cell types by detecting *Sox1* expression in vivo using a reporter system. For example, such a reporter system may comprise a readily identifiable marker under the control of a *Sox1* activated expression system. Fluorescent markers, which can be detected and sorted by FACS, are preferred. Especially preferred are GFP and luciferase.

Alternatively, an *in vivo* construct expressing a reporter may be placed under the control of the *Sox1* control sequences themselves. These sequences are activated at the same time as *Sox1* expression is activated, and therefore mark the transition into the neural pathway with the same accuracy as *Sox1*. Advantageously, the *Sox1* control sequences used are human *Sox1* control sequences. Preferably, they comprise nucleotides 1 to 60 of SEQ. ID. No. 3.

In general, reporter constructs useful for detecting neural cells by expression of a reporter gene may be constructed the general teaching of Sambrook et al (1989). Typically, constructs according to the invention comprise a promoter by *Sox1*, and a coding sequence encoding the desired reporter constructs, for example of GFP or luciferase. Vectors encoding GFP and luciferase are known in the art and available commercially.

SOX proteins bind to a sequence motif (A/T A/T CAA A/T G) (SEQ. ID. No. 6) with high affinity. Accordingly, constructs according to the invention advantageously comprise the above-recited motif, or a functional equivalent thereof, operably linked to a gene encoding a selectable marker.

When transfected into cells which are potentially express *Sox1*, constructs according to the invention will be activated specifically by *Sox1* expression. Therefore, the selectable marker will be expressed once the cell enters the neural differentiation pathway and *Sox1* expression is induced. This allows cells entering the neural differentiation pathway to be sorted by FACS.

In a still further aspect, the present invention relates to the transfection of pluripotent precursor cells, capable of differentiating into neural cells, with a vector expressing *Sox1*. By such means, pluripotent precursor cells may be induced to differentiate along the neural pathway, becoming precursor neurons capable of differentiating into a variety of neural tissues.

Herein, terms such as "transfection", "transformation" and the like are not intended to be significant, except to indicate that nucleic acid is transferred to a cell or organism in functional form. Such terms include various means of transferring nucleic acids to cells, including transfection with $CaPO_4$, electroporation, viral transduction, lipofection, delivery using liposomes and other delivery vehicles, biolistics and the like.

Suitable pluripotent precursor cells may be derived from a number of sources. For example, ES cells, such as human ES cells and cells derived from a Germ cells (EG cells) may be derived from embryonal tissue. Alternatively, pluripotent cells may be prepared by a retrodifferentiation, by the administration of growth factors or otherwise (see WO 96/23870), or by cloning, such as by nuclear transfer from an adult cell to a pluripotent cell such as an ovum.

Human stem cells of neural lineage may be isolated from human tissues directly. Alternatively, stem sells from non- human animals, such as rodents, may be used.

Neural stem cells may also be propagated in vitro, for example as described in Snyder *et al.* (1996) Clinical Neuroscience 3: 310-316, and Martinez-Serrano *et al.*,

(1996) Clinical Neuroscience 3:301-309. Moreover, pluripotent cell lines such as the N-Tera II cell line which are capable of differentiating into neural cells upon stimulation with agents such as retinoic acid are also responsive to *Sox1* stimulation.

The cDNA or genomic DNA encoding native or mutant SOX1, or a label under to control of *Sox1* sequences or a sequence transactivatable by SOX1, can be incorporated into vectors according too techniques known in the art. As used herein, vector (or plasmid) refers to discrete elements that are used to introduce heterologous DNA into cells for expression. Selection and use of such vehicles are well within the skill of the artisan. The vector components generally include, but are not limited to, one or more of the following: an origin of replication, one or more marker genes, an enhancer element, a promoter, a transcription termination sequence and a signal sequence.

Most expression vectors are shuttle vectors, i.e. they are capable of replication in at least one class of organisms but can be transfected into another class of organisms for expression. For example, a vector is cloned in E. coli and then the same vector is transfected into mammalian cells even though it is not capable of replicating independently of the host cell chromosome.

Advantageously, an expression and cloning vector may contain a selection gene, also referred to as selectable marker, other than that intended for marking *Sox1*-expressing cells. This gene may encode a protein necessary for the survival or growth of transformed host cells grown in a selective culture medium. Host cells not transformed with the vector containing the selection gene will not survive in the culture medium. Typical selection genes encode proteins that confer resistance to antibiotics and other toxins, e.g. ampicillin, neomycin, methotrexate or tetracycline, complement auxotrophic deficiencies, or supply critical nutrients not available from complex media.

Since the replication of vectors is conveniently done in E. coli, an E. coli genetic marker and an E. coli origin of replication are advantageously included. These can be obtained from E. coli plasmids, such as pBR322, Bluescript© vector or a pUC plasmid, e.g. pUC18 or pUC19, which contain both E. coli replication origin and E. coli genetic marker conferring resistance to antibiotics, such as ampicillin.

Expression vectors usually contain a promoter that is recognised by the host organism and is operably linked to SOX1, or label-encoding, nucleic acid. Such a promoter may be inducible by factors which induce *Sox1*, or by *Sox1* itself. The promoters are operably linked to DNA encoding SOX1 by removing the promoter from the source DNA and inserting the isolated promoter sequence into the vector. Both the native SOX1 promoter sequence and many heterologous promoters may be used to direct amplification and/or expression of SOX1 DNA. The term "operably linked" refers to a juxtaposition wherein the components described are in a relationship permitting them to function in their intended manner. A control sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

Control sequences, comprising a promoter and optionally enhancer(s), may be derived from the human or other *Sox1* genes. Alternatively, any suitable promoter may be used, when placed under the control of a SOX1-inducible element. In such a construct, the promoter selected should have a low residual level of activity, such as to minimise expression of the label in the absence of *Sox1* expression.

The vectors may also contain sequences necessary for the termination of transcription and for stabilising the mRNA. Such sequences are commonly available from the 5' and 3' untranslated regions of eukaryotic or viral DNAs or cDNAs. These regions contain nucleotide segments transcribed as polyadenylated fragments in the untranslated portion of the mRNA encoding SOX1 or the label.

An expression vector includes any vector capable of expressing SOX1 or label-encoding nucleic acids that are operatively linked with regulatory sequences, such as promoter regions, that are capable of expression of such DNAs. Thus, an expression vector refers to a recombinant DNA or RNA construct, such as a plasmid, a phage, recombinant virus or other vector, that upon introduction into an appropriate host cell, results in expression of the cloned DNA. Appropriate expression vectors are well known to those with ordinary skill in the art and include those that are replicable in eukaryotic and/or prokaryotic cells and those that remain episomal or those which integrate into the host cell genome. For example, DNAs encoding SOX1 may be inserted into a vector suitable for expression of cDNAs in mammalian cells, e.g. a CMV enhancer-based vector such as pEVRF (Matthias, et al., (1989) NAR 17, 6418).

Particularly useful for practising the present invention are expression vectors that provide for the transient expression of DNA encoding SOX1 or a label in mammalian cells. Transient expression usually involves the use of an expression vector that is able to replicate efficiently in a host cell, such that the host cell accumulates many copies of the expression vector, and, in turn, synthesises high levels of SOX1or a label. For the purposes of the present invention, transient expression systems are useful e.g. for identifying SOX1 expressing cells or for inducing a pluripotent cell to differentiate.

Construction of vectors according to the invention employs conventional techniques, for example as described in Sambrook *et al.*, 1989. Isolated plasmids or DNA fragments are cleaved, tailored, and religated in the form desired to generate the plasmids required. If desired, analysis to confirm correct sequences in the constructed plasmids is performed in a known fashion. Suitable methods for constructing expression vectors, preparing in vitro transcripts, introducing DNA into host cells, and performing analyses for assessing gene expression and function are known to those skilled in the art. Gene presence, amplification and/or expression may be measured in a sample directly, for example, by conventional

Southern blotting, Northern blotting to quantitate the transcription of mRNA, dot blotting (DNA or RNA analysis), or in situ hybridisation, using an appropriately labelled probe which may be based on a sequence provided herein. Those skilled in the art will readily envisage how these methods may be modified, if desired.

The invention is described, for the purpose of illustration only, in the following examples.

MATERIAL AND METHODS

Manufacture of SOX1 polyclonal antibodies: A 622bp HincII fragment encoding sequences C-terminal of the HMG box of SOX1 (207 a.a.) is fused in frame to the bacterial GST gene in the construct pGEX3X. Fusion protein is induced and purified as described by Smith and Johnson (1988) Gene 67:31-40. rabbits are treated with a course of injections as recommended by Smith and Johnson (1988): each injection contains 250μg of fusion protein. Two final bleeds, FB43 and FB44, are obtained from the rabbits prior to the preparation of polyclonal sera.

Immunocytochemistry: Embryos, P19 cells and neural plate explants are examined using standard techniques (Placzek *et al.*, (1993) Development 117:205-218). Antibodies are used at the following dilutions: anti-SOX1 PAb (1:500); K2 anti-HNF3β MAb (1:40); 6G3 anti-FP3 MAb (1:10); anti-3A10 MAb (1:10); anti-2H3(Neurofilament-160) MAb (1:10); 4D5 anti-Islet-1 MAb (1:1000); anti-SSEA1 MAb (1:80) (Hybridoma Bank); anti-NESTINE MAb (1:10) (Hybridoma Bank); anti-BrDU MAb (1:500) (Sigma); Appropriate secondary antibodies (TAGO and Sigma) are conjugated to fluorescein isothiocyanate (FITC), Cy2 or Cy3.

BrDU analysis: Pregnant mice are injected intraperitoneally with 50μg/g of body weight of 5-bromo-2deoxyuridine (BrDU) (Sigma) in 09.% NaCl and sacrificed two hours after injection. Embryos are fixed and sectioned as described above. The slides are washed twice in PBS, and incubated in 0.2% HCl at 37°C for 30 minutes, then rinsed thoroughly with PBS, followed by three rinses with PBS/0.1%

Trinton/1% heat inactivated goat serum (P-T-G). Monoclonal anti-BrDU (1:500 dilution in P-T-G) is applied to the sections and incubated at 4°C overnight. Sequential sections are incubated in SOX1 antibody (1:500 dilution in P-T-G) at 4°C overnight. The slides are washed twice in P-T-G, then incubated in the appropriate secondary antibody for 30 minutes at room temperature, washed with P-T-G and mounted.

P19 cell cultured and retinoic acid treatment: P19 cells are cultured as previously described (Rudnichy and McBurney, 1987). To induce differentiation, cells are allowed to aggregate in bacterial grade petri dishes alone, in the presence of 1μM retinoic acid or in the presence of 1μM retinoic acid or in the presence of 5mM IPTG. After 4 days of aggregation in the presence of inducing agents, cells are plated on tissue culture chamber slides. The cells are allowed to adhere and grow for 4-5 days, with media changes every 24 hours. For immunoflurescence, cells are grown on tissue culture chamber slides coated with 0.1% gelatin, washed once with PBS, fixed at room temperature in 1x MEMFA for 1 hour, washed in P-T-G twice; then stained with appropriate antibody.

Cell counting analysis: For cell counting experiments P19 transfectant cell lines are induced to differentiate, plated on gelatine coated slides, fixed at room temperature in 1xMEMFA for one hour at day 6-8 for neurons. Cells are stained with Neurofilament (2H3) antibody and photographed using an Olympus fluorescence microscope. Cell counts are expressed as percentages of total cells in a field. Eight fields from two different experiments are counted for each P19 clone.

Plasmids and transfection: To construct the SOX1 expression vector, pRSVop*Sox1*, the POP113CAT operator vector (Stratagene) is digested with Not1, end-filled Kpn/Stu (position 431-1694) fragment of the *Sox1* cDNA. The P3'SS, eukaryotic Lac repressor expressing vector (obtained from Stratagene) is transfected into P19 cells by lipofection. Stable transformants are selected in 250 μg/ml of hygromycin. Expanded clones (250) are isolated and examined for expression of

the Lac repressor by indirect immunofluorescence with anti-lac PAb (Stratagene). Four cell lines are isolated (P3'SS-10, 13, 22 and 47) which show ubiquitous and constitutive expression of the Lac repressor. P3'SS-10 is chosen for the subsequent experiments. P3'SS-10 is then transfected with pRSVop*Sox1* by lipofection. Stable clones are selected using 500μg/ml G481. 250 clones are expanded and analysed for inducible *Sox1* expression by RNase protection and immunocytochemistry with SOX1 antibody.

RNase protection assays: Total RNA is prepared from P19 cells and RNase protection assays are carried out using 5μg of P19 cell RAN as described by Capel *et al.*, (1993) Cell 73:1019-1030. Anti-sense labelled probes are derived from the 396 bp SmaI-BspH1 fragment (position 1467-1863) of the *Sox1* cDNA, a 215bp BsaI exon 4 specific fragment of *Wnt1* cDNA, a PvuII digest of the *Mash1* cDNA (Johnson *et al.*, (1992) Development 114:75-87) and a NotI digest of SAP D cDNA is used a loading control (Dresser *et al.*, (1995) Hum. Mol. Genet. 4:1613-1618).

RT-PCR: Total RNA is prepared from P19 cells as described by Capel *et al.*, (1993). Reserve transcription, PCR reaction, and priming is performed as described by Okabe *et al.*, (1996).

Rat lateral neural plate explants: Lateral neural plates (LNP) are isolated from days 8.5-9.0 rat embryos from prospective hindbrain and spinal cord regions as previously described (Placzek *et al.*, 1993). Notochord explants are dissected from HH stage 608 chick embryos as previously described (Placzek *et al.*, 1993). Explants are embedded in collagen and cultured (Placzek *et al.*, 1993) for 24, 48 and 96 hours. Purified rat SHH-N (Ericson *et al.*, (1996) Cell 87:661-673) is added to cultures at concentrations within the effective ranges used in other assays (Ericson *et al.*, 1996)

EXAMPLE 1

SOX1 IS EXPRESSED DURING EARLY NEURAL DEVELOPMENT

SOX1 expression during mouse and rat neurulation is analysed using a rabbit polyclonal antibody against the SOX1 C-terminal region. In the mouse, expression of SOX1 is first detected at 7.5 days post coitum (*dpc*) in the anterior half of the late-streak egg cylinder. Cross-sections through the embryo at this stage reveal expression in columnar ectodermal cells, which appear to define the neural plate, while cells located more laterally are negative. Thus, SOX1 expression at this stage is specific to the neural plate. SOX1 is maintained in all neuroepitheial cells along the entire anteroposterior axis as the neural pate bends (8.0-8.5 dpc, as shown in cross-sections of a 2 somite mouse embryos where *Sox1* expression is limited to neural folds) and fuses to form the neural tube (9.0-9.5 dpc, where *Sox1* labelling is seen to be restricted to the neural tube in cross-sections of 10-12 somite mouse embryos). The pattern of expression of SOX1 in the rat is similar to that in the mouse. The expression of SOX1 throughout the neural plate and early neural tube implies a similarity amongst these cells.

After neural tube closure, neuroepithelial cells begin to differentiate into defined classes of neurons at specific dorsoventral (D/V) positions within the spinal cord (Altman and Bayer (1984) Adv. Anat. Embryol. Cell Biol. 85:32-46; Tanabe and Jessell, (1996) Science 274:1115-1123). As development proceeds, *Sox1* is downregulated in a stereotyped manner in cells alone D/V axis of the neural tube. In the spinal cord, expressions first downregulated in cells that occupy the ventral midline (cross-sections of the thoracic region of 20 somite mouse embryos reveal a lack of SOX1 staining in this area), then the ventral motor horns (corresponding lack of staining being visible in cross section of 30-35 somite embryos) and subsequently the dorsal regions. These regions appear to correlate with floor plate, motor neurons and sensory relay interneurons, respectively.

To ascertain this a series of antibody double-labelling experiments are performed in rat embryos. The SOX1 antibody is used in combination with a panel of antigenic markers which identify cells of the floor plate and mature neurons (Neurofilament (NF-1): labelled with contrasting colour markers and visualised in an E11 rat embryo). Expression of SOX1 and expression of these markers is almost entirely mutually exclusive. In the ventral spinal cord or the 10.0-12.0 *dpc* mouse embryo, SOX1 expression is maintained only in 'region X' (Yamada *et al.*, (1991) Cell 64:635-647), as revealed by immunolabelling of two streams of cells located between the differentiated floor plate and ventral motor horns in 30-35 somite embryos. Eventually, by 13.5 *dpc*, SOX1 expression is restricted to a thin ventricular zone in the CNS. SOX1 expression in to detected in the peripheral nervous system (PNS). These expression profiles suggest that SOX1 is expressed by early neural cells in the CNS and is downregulated in the developing neural tube coincident with neural differentiation.

EXAMPLE 2

SOX1 MARKS PROLIFERATION CELLS WITHIN THE EMBRYONIC NEURAL TUBE

The uniform expression of SOX1 in the neural plate and early neural tube followed by its down regulation along the D/V axis and restriction to the ventricular zone is reminiscent of the pattern of cell proliferation in the developing central nervous system (Sauer, (1935) J. Comp. Neurol. 62:377-405; Fujita, (1963) J. Comp. Neurol. 120: 37-42; Altman and Bayer, 1984). In the neural plate and early neural tube, proliferating progenitor cells are organised in a pseudostratified epithelium in which the processes of these cells extend from the inner luminal to the outer mantle surface. At later stages the neural tube becomes progressively thicker and can be divided into different zones. The proliferating CNS progenitors are largely restricted to the inner ventricular zone (VZ) around the lumen. They begin to migrate away from the lumen while in S-phase, and after completing their final

mitosis, migrate to the outer layer, the marginal zone (MZ). In the 10.5 dpc mouse embryo, SOX1 expression is detected, using an anti-SOX1 antibody, throughout the pseudostratified epithelium of the posterior neural tube and is restricted to the ventricular zone in more mature anterior region of the neural tube. In order to evaluate the relationship between SOX1 expression and proliferating CNS cells are directly assayed proliferation by monitoring the incorporation of bromodeoxyuridine (BrDU) with an anti-BrDU antibody. Pregnant mouse females at 10.5 *dpc* are injected with BrDU two hours prior to dissection to detect proliferating cells. Embryos are then fixed, sectioned and double-labelled for BrDU incorporation and SOX1 expression. Similar to SOX1 expressing cells, those that incorporate BrDU are found throughout the posterior neural tube in 10.5 *dpc* mouse embryos and lie in the ventricular zone of the anterior neural tube. All cells that incorporate BrDU also express SOX1. SOX1-positive cells that do not incorporate BrDU are restricted to the luminar surface of the ventricular zone. In contrast, no SOX1 nor BrDU-positive cells are detected in the outer marginal zone. These results show that SOX1 is expressed in dividing neuroepithelial cells within the embryonic CNS.

EXAMPLE 3

SOX1 IS DOWNREGULATED IN COMMITED CELLS

The mutual exclusion of SOX1 and markers of committed differentiated cells such as Islet1 (Pfaff *et al.*, (1996) Cell 84:1-20) raises the possibility that the downregulation of SOX1 may be a pre-requisite step for the differentiation in neural plate explants *in vitro*. Isolated neural plates explants are cultured with known inducers of ventral neural cells, namely the notochord and purified Sonic Hedgehog protein. The expression of SOX1 and incorporation of BrDU is then compared to the expression of three markers of ventral cells, Islet1, FP3 and HNF3β. Consistent with our observations *in vivo* both the expression of SOX1 and Islet1 as well as SOX1 and FP3 is mutually exclusive in neural plate explants cultured adjacent to notochord (n=8) or in the presence of purified Sonic Hedgehog protein as seen in E9 rat neural plate tissue cultured with Sonic Hedgehog protein for 48

hours and stained with anti-SOX1 and anti-Islet1 antibodies. Similarly, the incorporation of both BrDU and Islet1 as well as BrDU and FP3 (detected using an anti-FP3 antibody) is mutually exclusive. In contrast, the domain of expression of HNF3β is found to extend beyond that of FP3 and into the region of BrDU positive cells.

To determine whether a similar population of cells could be detected *in vivo* embryos are analysed, and for co-expression of FP3 and HNF3β and for co-expression of BrDU and HNF3β. We find that medial floor plate cells co-express HNF3β and FP3 but do not incorporate BrDU, whereas lateral floor plate cells express only HNF3β and incorporate BrDU. HNF3β thus provides a marker for cells that are mitotically active but have begun to differentiate.

These cells, occupying the medial regions of the floor plate, express HNF3β but not SOX1. In contrast cells occupying lateral regions of the floor plate co-express HNF3β and SOX1. These observations, together with the mutually exclusive expression of SOX1 with Islet 1 and FP3 in ventral neural cells provide evidence that SOX1 is downregulated as cells exit mitosis and not at the onset of cell differentiation.

EXAMPLE 4

SOX1 EXPRESSION IS ASSOCIATED WITH NEURAL DIFFERENTIATION

Neural induction is accompanied by the onset of new gene expression which in turn enables the formation of neural rather than epidermal tissue. The early and apparently uniform expression of SOX1 in neural cells, together with observations that *Sox* genes may affect cell lineage decisions, raises the possibility that SOX1 expression is an early response to neural inducing signals and that its expression may be involved in directing cells towards a neural fate. To address whether SOX1 plays a role in establishing neural fate in response to A P19 cell culture system is

used as an *in vivo* model system in which to analyse SOX1 expression and the effects of its misexpression.

P19 cells are an embryonal carcinoma cell line with the ability to differentiate into all three germ layers (McBurney, (1993) Int. J. dev. Biol. 37:135-140). In the undifferentiated state P19 cells morphologically resemble an uncommitted primitive ectodermal cell and express the cell surface antigen SSEA-1. These cells have a very low rate of spontaneous differentiation when grown in a monolayer in the absence of chemical inducers. P19 cells grown as aggregates, however, differentiate partially into endodermal cells. Furthermore, with the addition of retinoic acid, aggregated P19 cells differentiate into neuroepithelial-like cells (Jone-Villeneuve *et al.*, (1982) J. Cell. Biol. 94:253-262). These express neuroepithelial markers such as NCAM, intermediate filament NESRIN, MASHI (Johnson *et al.*, 1992) and WNT1 (St. Arnaud *et al.*, (1989) Oncogene 4:1077-1080). When plated onto a substrate, about 15% of these cells differentiate into mature neurons expressing Neurofilament. Thus, in this *in vitro* model system retinoic acid acts as a "neural inducer".

Initially, the expression of *Sox1* in P19 cells is examined by both RNase protection and immunocytochemistry. The features of *Sox1* expression in P19 cells are similar to those observed in prospective neural tissue in vivo. *Sox1* mRNA and protein can not be detected in undifferentiated P19 cells which express the cell-surface antigen SSEA1 when analysed using anti-SOX1 and anti-SSEA antibodies, and by RNaše protection. Similarly, when P19 cells are differentiated as aggregates without the addition of chemical inducers, SOX1 is not expressed as determined by RNase protection. In contrast, SOX1 is rapidly induced during neural differentiation when aggregated P19 cells are differentiated in the presence of retinoic acid. *Sox1* thus behaves similarly to other neuroepithelial markers such as *Mash 1* and *Wnt 1*, the transcripts of which are detected in retinoic acid-treated P19 cells by RNase protection.

When retinoic acid-treated P19 cell aggregates are plated onto tissue culture substrate, about 15% of the cells differentiate into mature process-bearing, Neurofilament-expressing neurons. Double-label immunofluorescence is used to simultaneously detect SOX1 and Neurofilament, to examine the expression of SOX1 in P19 cells displaying a fully differentiated neuronal morphology. SOX1 immunoreactivity is not detected in process-bearing Neurofilament-positive neurons. Thus, as in vivo, SOX1 is expressed by P19 cells when they first assume a neural fate but it is then downregulated with their differentiation.

EXAMPLE 5
USE OF SOX1 TO DIRECT CELLS TO A NEURAL FATE

The previous data suggest that in P19 cells, as in vivo, SOX1 expression is induced at a time when neuroepithelial cells begin to differentiate. If SOX1 plays a role in directing cells towards the neural fate, expression of SOX1 in P19 cells may be able to substitute for retinoic acid to initiate neural differentiation. Endogenous SOX1 is accordingly activated in P19 cells using an inducible eukaryotic lac repressor-operator expression system. To establish this system a clonal line of P19 cells is generated which constitutively and ubiquitously expresses the lac repressor. This parent line (P3'SS-10) is transfected with pRSVop*Sox1*, a vector containing the *Sox1* cDNA under the regulation of an inducible RSV promoter and stable lines are established. In the uninduced state, without the addition of isopropyl-β-d-thiogalactase (IPTG) these lines express high levels of the lac repressor that binds to operon sites upstream of the RSV promoter and thus blocks transcription of *Sox1*. Upon addition of IPTG a conformational change occurs, decreasing the affinity of the repressor and resulting in the activation of pRSVop*Sox1*. Approximately 250 clones of transfectants are isolated in the repressed state. Using RNase protection and immunocytochemistry assays three clones are selected (708-13, 708-16 and 708-21) that express high levels of RSVop*Sox1* in response to IPTG.

The pluripotentiality of these clones is not compromised by the transfection and selection. All three lines express SSEA1 in the uninduced state. Furthermore, when aggregated in retinoic acid the uninduced clones initiate expression of endogenous *Sox1* and differentiate into mature Neurofilament-expressing neurons after plating, in a manner similar to wild-type P19 untransfected cells.

In order to address whether expression of SOX1 can initiate neural differentiation and thereby substitute for the requirement of retinoic acid, it is determined whether the transient exposure of P19 aggregates to retinoic acid can be replaced by a transient induction of RSVop*Sox1*, through addition of IPTG. Wildtype P19 cells and transfected P19 clones (708-13, 708-16 and 708-21) are cultured as aggregates for 96 hours with or without the addition of IPTG. After 96 hours RNA is isolated from half of the aggregates for RNase protection and/or RT-PCR assays. The remaining aggregates are plated onto tissue culture substrate, allowed to differentiate for three days without further addition of IPTG and then scored for the expression of a panel of neuroepithelial and neuronal markers by immunocytochemistry. These conditions are the same as those used for retinoic acid-induced differentiation of wildtype P19 cells. After 96 hours the clones induced to express RSVop*Sox1* with IPTG express endogenous *Sox1* and *Mash1*. The expression of these two neuroepithelial markers is similar to that seen in wildtype cells induced with retinoic acid. In addition the IPTG induced clones expressed NESTIN and *Hoxa7* (Mahon et al., (1988) Development (Suppl.) 187-195). Further differentiation of the transiently-induced clones on substrate showed the presence of mature neurons as demonstrated by Neurofilament-positive, 3A10-positive and Islet1-positive cells. All three clones 708-13, 708-16 and 708-21 differentiate in this matter although the number of mature neurons produced is variable. The number of differentiated neurons formed in the IPTG induced clones is estimated by determining the number of Neurofilament-positive cells in a given field of cells. The number of neurons ranges from 6-8% for clone 708-13, 15-20% for clone 708-16 and 20-25% for clone 708-21. The latter two clones show uniform and ubiquitous induction of SOX1 expression whereas expression in clone 708-13 is

not in all cells. In addition, the transiently induced clones generate GFAP-positive cells indicating glial cell differentiation. None of these markers is detected in wildtype P19 cells cultured in the presence of IPTG or in clones 708-13, 708-16, and 708-21 cultured in the absence of IPTG. The expression of SOX1, both in vivo and in vitro, is mutually exclusive with mature neuronal markers such as Neurofilament and Islet1. To examine SOX1 expression in the mature neurons generated in the transiently-induced clones, double-label immunoflourescence is used to simultaneously detect SOX1 and Neurofilament. No SOX1 expression could be detected in cells positive for Neurofilament in these cultures.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: MEDICAL RESEARCH COUNCIL
(B) STREET: 20 PARK CRESCENT
(C) CITY: LONDON
(E) COUNTRY: UK
(F) POSTAL CODE (ZIP): W1N 4AL

(ii) TITLE OF INVENTION: NEURONAL STEM CELL GENE

(iii) NUMBER OF SEQUENCES: 6

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2312 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Gallus gallus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CCGCAGAGCG CGGCAGGACG GGCACACGCC GGGCGCAGCA CCGCGAAGCA CCGCGCAGCC 60

CCGCGCAGCC CCGCACCTGT TTGCGCGCCC CGCGCCCGGA GCGGCCCCCG GCAGCGGGAG 120

GACGCCGGCA GCGCCGCCGC CGCCGCTCCT CGCATGTGCG GTGCCTCCCC GCCGCCCGGC 180

GCCGGAGGGA AGTGAGGAAG CCCCGTGAAT GTACAGCATG ATGATGGAGA CGGACTTGCA 240

CTCGCCCGGC GGAGCCCCGG CGCCCGGCGG CGGCCTCTCG GGGCAGAGCG GCGCGGGCGG 300

CGGCGGCGGC GGCGGCGGCG GCGGCGGGGG CAAAGCGGGG CAGGACCGCG TGAAGCGCCC 360

CATGAACGCC TTCATGGTGT GGTCGCGGGG GCAGCGGCGG AAGATGGCCC AGGAGAATCC 420

CAAGATGCAC AACTCGGAGA TCAGCAAGCG GCTGGGCGCC GAGTGGAAGG TGATGTCGGA 480

GGCCGAGAAG CGGCCTTTCA TCGACGAGGC GAAGCGGCTG CGGGCGCTGC ACATGAAGGA 540

GCACCCGGAT TATAAATACC GGCCCCGGCG GAAGACCAAG ACGCTGCTCA AGAAGGACAA 600

GTACTCGCTG GCCGGAGGGC TGCTGGGCGC CGGCCCGGCC GCGGGCGGCC CTCCCGCCGT 660

CGGCGTGGGC ATGGGCGTCG GCGTGATCCC CGGCGGAGTC GGGCAGCGGC TGGAGAGCCC 720

CGGCGGGGCG GCGGGCGGCG GCTACGCGCA CATGAACGGG TGGGCCAACG GCGCCTACCC 780

GGGCTCGGTG GCGGCGGCGG CGGCGGCGGC GGCGATGATG CAGGAGGCGC AGCTCGCCTA 840

CGGGCAGCAC CCGGGCGGCG GGGGGCACCC GCACCACCCG CACCCGCACC ACCCGCACCA 900

CCCGCACAAC CCGCAGCCCA TGCACCGCTA CGACATGGGC GCGCTGCAGT ACAGCCCCAT 960

CTCCAACTCG CAGGGCTACG TGAGCGCCTC GCCCTCGGGC TACGGCGCGC TGCCCTACGG 1020

CTCGCAGCCC CACCAGAACT CGGCGGCCGC GGCGGCGGCG GCGGCGGCGG CGGCGGCCGC 1080

CTCGTCGGGC GCGCTGGGCG CGCTGGGCTC GCTCGTCAAG TCGGAGCCCA GCGTGAGCCC 1140

GCCCGTCACC TCGCACTCGC GGGCCCCGTG CCCCGGGGAC CTGCGGGAGA TGATCAGTAT 1200

GTACTTACCG GGCGGCGAGG GAGGCGACCC GGCGGCCGCC GCCGCCCAGA GCCGCCTGCA 1260

CTCCCTGCCC CAGCACTACC AGAGCGCCAG CACGGGGGTC AACGGCACCG TCCCCTTGAC 1320

GCACATCTGA GCGGCCCCGG AGCGGCCCCG GAGCGGCGCG GAGGGCCCCG GCCCGGGCCC 1380

CGCAGGACTG CGGCCCCGCC GCCGCCCCGC GCCCGCCGCC CCCCTTCGTT TTTGCCTTTC 1440

ATTCGGCTCC TTCCCGCCCT CCCCCTCCCT CCTTCCTTTT TTTGTTTTGT TTTGTTTTGT 1500

TTTTCTTTTC TTCCTTTTTG TACAGAAATG TTTTGATGTT CTTGTAATAA TAATAAATAA 1560

TAATAATAAT AATAACGAGA GAGAGAGAAA AAGAAGGTAA CGGTGGCTTT ACTGACCTTT 1620

TTGTTTTTAG GAGGACCAGA TCCCGGGACT AGTTTTAGAC TGAACTTCTG TGTTTTATCG 1680

AGACTTTTTG TACAGTATTT ATCATTCACC CCAGAGACAC AGAGCGTTTA TTTGCAAAAG 1740

AGGAGAGAGA GAGGATTAAA AAACAAAAAC AAACAAACAA ACAAACAAAA AAAGACGGCG 1800

ACGAAAAGAC AAAACCATCG CCGCTGACAC CCAAAGTTCG GGCGGGGCCA ACTTTCGGGC 1860

TGCGCTTCGC CCCGCACCGC CTCACTGCAA ACGGAGCCGA CGGGGAGCGG TGCTCGTTCC 1920

TTCCTCGCAC ACCCCAAAAC AGCACCACGA GTTTCCGTAG ATGTTCTCGC GCTTTTCCTT 1980

TTTGGTTGGG TTATTTCGGC TGCTTTATTT ATACAACTTT TTCTTCTTCT TCCTTTCTTC 2040

CCGAGGTTGC AACGTTTGCT TGATTTTTAT TTTATTTTAT TTTTTTTTCT GGGTTATGTG 2100

AAACTTTACT GTATCTGCAT CATTTCGGTT TGTTTTCCTC CCCCCCCCCC CCTTTTTTTT 2160

TTTTTTTACA TTTTTTTGTA TCATCTCGTG TAAATGCATT GTGAAATAAT TTTTATCTAG 2220

GCGTGGCGAG GGAACCCAGA CTGTACATAG TTTACTAAAA AGCCTTTCTG CTAAACAGAA 2280

ACCCGAAGGA TGCGTTCCAT TTTGAGTTAA AT 2312

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2376 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mus musculus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

ACAGGAACGG AGACTTCGAG CCGAGAAGAG GAGGCAGCGA ACCCTGCGTC GGGCCCAGGG 60

GCACCGCTTC AGACCCAGAA AGTGGAGCCT CAACTTGGCC ACGACTGCAC CTGTTTGCAC 120

AGTTCAGCCC TGAGTGACCG GACGGCAGCA ATCAACCTGG CCATCGGCCT CTTTGGCAAG 180

TGGTTTGTGC ACCGGGAGAA ACTTTCCACC TGCGAGCTGG ACCCGCGCTA AGTGCGTGTG 240

CTTTTGCCTC TTTTTTGTTG TTGTTGTTGT TGTGGCCTCC ACCCAACCCC CTTCTCTCCG 300

CTAGGCACCC ACCGCACACA CACCCCCCCC CCCAGTCTCT CTGGGCTGAT CCTCTCTCCA 360

CCCACCCACC CCCACCCGGC CGTCTATGCT CCAGGCCCTC TCTTTGCGGT ACCGGTGAAC 420

CCGCTAGCCG CCCAGATGTA CAGCATGATG ATGGAGACCG ACCTGCACTC GCCCGGCGGC 480

GCCCAGGCGC CCACGAACCT CTCGGGCCCG GCCGGGGCGC GCGGGGGCGG CGGTGGGGGC 540

GGGGGCGGCG GCGGCGGCGG GGGCACCAAG GCCAACCAGG ATCGGGTCAA GCGGCCCATG 600

AACGCCTTCA TGGTGTGGTC CCGCGGACAG CGGCGCAAGA TGGCCCAGGA AAACCCCAAG 660

ATGCACAACT CGGAGATCAG CAAGCGCCTC GGGGCCGAGT GGAAGGTCAT GTCCGAGGCC 720

GAGAAGCGGC CGTTCATCGA CGAGGCCAAG AGACTGCGCG CGCTGCACAT GAAGGAACAC 780

CCGGATTACA AGTACCGGCC GCGCCGCAAG ACCAAGACGC TGCTCAAGAA GGACAAGTAC 840

TCGCTGGCCG GCGGGCTGCT AGCGGCCGGC GCGGGTGGCG GCGGCGCGGC CGTGGCCATG 900

GGTGTGGGCG TGGGCGTCGG GGCGGCGGCG GTGGGCCAGC GCTTGGAGAG CCCAGGCGGC 960

GCGGCGGGAG GAGGCTACGC GCATGTCAAC GGCTGGGCTA ACGGCGCCTA CCCCGGCTCG 1020

GTGGCCGCGG CGGCTGCGGC CGCGGCCATG ATGCAGGAGG CACAGCTGGC CTACGGGCAG 1080

CACCCAGGCG CGGGCGGCCG GCACCCGCAC GCACACCCGG CGCACCCGCA CCCGCACCAT 1140

CCGCACGCGC ATCCTCACAA CCCGCAGCCC ATGCACCGCT ACGACATGGG CGCGCTGCAG 1200

TACAGCCCCA TCTCCAACTC TCAGGGCTAC ATGAGCGCGT CGCCTTCGGG CTACGGCGGC 1260

ATCCCTTACG GCGCCGCGGC CGCCGCCGCC GCCGCTGCGG GCGGCGCGCA CCAGAACTCG 1320

GCGGTGGCGG CAGCGGCAGC CGCGGCAGCC GCGTCGTCGG GGGCCCTGGG CGCCCTCGGA 1380

TCTCTGGTCA AGTCGGAGCC GAGCGGCAGT CCGCCGGCCC CGGCTCACTC ACGGGCACCG 1440

TGTCCCGGGG ACCTGCGCGA GATGATCAGC ATGTACCTGC CGGCCGGCGA GGGTGGCGAC 1500

CCGGCGGCGG CAGCGGCTGC GGCGGCCCAA AGCCGGCTGC ACTCGCTGCC ACAGCACTAC 1560

CAGGGCGCGG GCGCGGGCGT CAACGGCACG GTGCCCCTGA CGCACATCTA GCGCCGCGGG 1620

GACGCCGGGG ACACTGCGGC TTAAGGCCGG CGCCCCGGCG ACGAAGAGCG AGGCCTGCGC 1680

CCCAGCCTCC AGAGCCCGAC TTTGTACCGA GGTCCCCGCG CTCTCGATAA AAGGCCGCTC 1740

TGGAGAGCCG AGCGCCAGGT GACATCTGCC CCCATCACCT TCCCCAGGAC TCCGAGGCGC 1800

TGACACCAGA CTGGCCTCTT AGACTGAACT TTGGTGTTTT CATGAGACCT TTTGTACAGT 1860

ATTTATCGTC CGCAGAGGAG GCACACAGCG TTTTCTCGGC TTCGGAGGAC AAAAGACAAA 1920

AACCCAGCGA GGCGATGCCA ACTTTTGTAT GACTGCCGGC TCTGTAACTT TTTCCGGGGT 1980

TTACTTCCCG CCAGCTCTTC TGCCTGAGGC CGAGTGACGG ACCTCGAGCC CTTCTCACTT 2040

GTTATAAATC TAAGTAAGGC AGATCCAAAC ATTTACAAGT TTTTTGTAGT TGTTACCGCT 2100

CTTTTGGGTT GGTTGGTTAA TTTATACCGC AATCCCCTCT CAGACGGTGG AGTTATATTC 2160

TGGGTTTTGT AAATCTCTGT ATCCGAGCAT TTCCAATTTT TTGTTTTGTT TTGTATTATT 2220

TCTTGTAAAT GCGTTGTGAC ATTTTTATTT TAGGCGTTGC GATACGGGGG GAAGAGGAGT 2280

CGGATGTTGT ACATAGCCTG CAAGTCTTTC ATCTAAAAGC AAAAACAAAG AGAGATACCC 2340

CCAAAATGCA TCAAATTTGA ACAATACATT TAAGAG 2376

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1542 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA to mRNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Homo sapiens

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION:60..1223

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

```
CCGGCCGTCT ATCTCCAGGC CCTCTCCTCG CGGTGCCGGT GAACCCGCCA GCCGCCCCG        59

ATG TAC AGC ATG ATG ATG GAG ACC GAC CTG CAC TCG CCC GGC GGC GCC        107
Met Tyr Ser Met Met Met Glu Thr Asp Leu His Ser Pro Gly Gly Ala
  1               5                  10                  15

CAG GCC CCC ACG AAC CTC TCG GGC CCC GCC GGG GCG GGC GGC GGC GGG        155
Gln Ala Pro Thr Asn Leu Ser Gly Pro Ala Gly Ala Gly Gly Gly Gly
             20                  25                  30

GGC GGA GGC GGG GGC GGC GGC GGC GGC GGG GGC GCC AAG GCC AAC CAG        203
Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Ala Lys Ala Asn Gln
         35                  40                  45

GAC CGG GTC AAA CGG CCC ATG AAC GCC TTC ATG GTG TGG TCC CGC GGG        251
Asp Arg Val Lys Arg Pro Met Asn Ala Phe Met Val Trp Ser Arg Gly
     50                  55                  60

CAG CGG CGC AAG ATG GCC CAG GAG AAC CCC AAG ATG CAC AAC TCG GAG        299
Gln Arg Arg Lys Met Ala Gln Glu Asn Pro Lys Met His Asn Ser Glu
 65                  70                  75                  80

ATC AGC AAG CGC CTG GGG GCC GAG TGG AAG GTC ATG TCC GAG GCC GAG        347
Ile Ser Lys Arg Leu Gly Ala Glu Trp Lys Val Met Ser Glu Ala Glu
                 85                  90                  95

AAG CGG CCG TTC ATC GAC GAG GCC AAG CGG CTG CGC GCG CTG CAC ATG        395
Lys Arg Pro Phe Ile Asp Glu Ala Lys Arg Leu Arg Ala Leu His Met
            100                 105                 110

AAG GAG CAC CCG GAT TAC AAG TAC CGG CCG CGC CGC AAG ACC AAG ACG        443
Lys Glu His Pro Asp Tyr Lys Tyr Arg Pro Arg Arg Lys Thr Lys Thr
        115                 120                 125
```

```
CTG CTC AAG AAG GAC AAG TAC TCG CTG GCC GGC GGG CTC CTG GCG GCC      491
Leu Leu Lys Lys Asp Lys Tyr Ser Leu Ala Gly Gly Leu Leu Ala Ala
    130                 135                 140

GGC GCG GGT GGC GGC GGC GCG GCT GTG GCC ATG GGC GTG GGC GTG GGC      539
Gly Ala Gly Gly Gly Gly Ala Ala Val Ala Met Gly Val Gly Val Gly
145                 150                 155                 160

GTG GGC GCG GCG CCC GTG GGC CAG CGC CTG GAG AGC CCA GGC GGC GCG      587
Val Gly Ala Ala Pro Val Gly Gln Arg Leu Glu Ser Pro Gly Gly Ala
                165                 170                 175

GCG GGC GGC GCG TAC GCG CAC GTC AAC GGC TGG GCC AAC GGC GCC TAC      635
Ala Gly Gly Ala Tyr Ala His Val Asn Gly Trp Ala Asn Gly Ala Tyr
            180                 185                 190

CCC GGC TCG GTG GCG GCC GCG GCG GCC GCC GCG GCC ATG ATG CAG GAG      683
Pro Gly Ser Val Ala Ala Ala Ala Ala Ala Ala Ala Met Met Gln Glu
        195                 200                 205

GCG CAG CTG GCC TAC GGG CAG CAC CCC GGC GCG GGC GGC GCG CAC CCG      731
Ala Gln Leu Ala Tyr Gly Gln His Pro Gly Ala Gly Gly Ala His Pro
    210                 215                 220

CAC CGC ACC CCG GCG CAC CCG CAC CCG CAC CAC CCG CAC GCG CAC CCG      779
His Arg Thr Pro Ala His Pro His Pro His His Pro His Ala His Pro
225                 230                 235                 240

CAC AAC CCG CAG CCC ATG CAC CGC TAC GAC ATG GGC GCG CTG CAG TAC      827
His Asn Pro Gln Pro Met His Arg Tyr Asp Met Gly Ala Leu Gln Tyr
                245                 250                 255

AGC CCC ATC TCC AAC TCG CAG GGC TAC ATG AGC GCG TCG CCC TCG GGC      875
Ser Pro Ile Ser Asn Ser Gln Gly Tyr Met Ser Ala Ser Pro Ser Gly
            260                 265                 270

TAC GGC GGC CTC CCC TAC GGC GCC GCG GCC GCC GCC GCC GCC GCG CAC      923
Tyr Gly Gly Leu Pro Tyr Gly Ala Ala Ala Ala Ala Ala Ala Ala His
        275                 280                 285

CAG AAC TCG GCC GTG GCG GCG GCG GCG GCG GCG GCG GCC GCG TCG TCG      971
Gln Asn Ser Ala Val Ala Ala Ala Ala Ala Ala Ala Ala Ala Ser Ser
    290                 295                 300

GGC GCC CTG GGC GCG CTG GGC TCT CTG GTG AAG TCG GAG CCC AGC GGC     1019
Gly Ala Leu Gly Ala Leu Gly Ser Leu Val Lys Ser Glu Pro Ser Gly
305                 310                 315                 320

AGC CCG CCC GCC CCA GCG CAC TCG CGG GCG CCG TGC CCC GGG GAC CTG     1067
Ser Pro Pro Ala Pro Ala His Ser Arg Ala Pro Cys Pro Gly Asp Leu
                325                 330                 335

CGC GAG ATG ATC AGC ATG TAC TTG CCC GCC GGC GAG GGG GGC GAC CCG     1115
Arg Glu Met Ile Ser Met Tyr Leu Pro Ala Gly Glu Gly Gly Asp Pro
            340                 345                 350
```

GCG GCG GCA GCA GCG GCC GCG GCG CAG AGC CGG CTG CAC TCG CTG CCG 1163
Ala Ala Ala Ala Ala Ala Ala Ala Gln Ser Arg Leu His Ser Leu Pro
355 360 365

CAG CAC TAC CAG GGC GCG GGC GCG GGC GTG AAC GGC ACG GTG CCC CTG 1211
Gln His Tyr Gln Gly Ala Gly Ala Gly Val Asn Gly Thr Val Pro Leu
370 375 380

ACG CAC ATC TAG CGCCTTCGGG ACGCCGGGGA CTCTGCGGCG GCGACCCACG 1263
Thr His Ile *
385

AGCTCGCGGC CCGCGCCCGG CTCCCGCCCC GCCCCGGCGC GGCGTGGCTT TTGTATCAGA 1323

CGTTCCCACA TTCTTGTCAA AAGGAAAATA CTGGAGACGA ACGCCGGGTG ACGCGTGTCC 1383

CCCACTCACC TTCCCCGGAG ACCCTGGCGA CCGCCGGGCG CTGACACCAG ACTTGGTTTA 1443

GACTGAACTT CGGTGTTTTC TTGAGACTTT TGTACAGTAT TTATCACCTA CGGAGGAAGC 1503

GGAAGCGTTT TCTTTGCTCG AGGGACAAAA AATGCAAAA 1542

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 388 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Met Tyr Ser Met Met Met Glu Thr Asp Leu His Ser Pro Gly Gly Ala
1 5 10 15

Gln Ala Pro Thr Asn Leu Ser Gly Pro Ala Gly Ala Gly Gly Gly Gly
20 25 30

Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly Ala Lys Ala Asn Gln
35 40 45

Asp Arg Val Lys Arg Pro Met Asn Ala Phe Met Val Trp Ser Arg Gly
50 55 60

Gln Arg Arg Lys Met Ala Gln Glu Asn Pro Lys Met His Asn Ser Glu
65 70 75 80

Ile Ser Lys Arg Leu Gly Ala Glu Trp Lys Val Met Ser Glu Ala Glu
85 90 95

Lys Arg Pro Phe Ile Asp Glu Ala Lys Arg Leu Arg Ala Leu His Met
100 105 110

Lys Glu His Pro Asp Tyr Lys Tyr Arg Pro Arg Arg Lys Thr Lys Thr

```
          115                 120                 125

Leu Leu Lys Lys Asp Lys Tyr Ser Leu Ala Gly Gly Leu Leu Ala Ala
    130                 135                 140

Gly Ala Gly Gly Gly Gly Ala Ala Val Ala Met Gly Val Gly Val Gly
145                 150                 155                 160

Val Gly Ala Ala Pro Val Gly Gln Arg Leu Glu Ser Pro Gly Gly Ala
                165                 170                 175

Ala Gly Gly Ala Tyr Ala His Val Asn Gly Trp Ala Asn Gly Ala Tyr
            180                 185                 190

Pro Gly Ser Val Ala Ala Ala Ala Ala Ala Ala Ala Met Met Gln Glu
        195                 200                 205

Ala Gln Leu Ala Tyr Gly Gln His Pro Gly Ala Gly Gly Ala His Pro
    210                 215                 220

His Arg Thr Pro Ala His Pro His Pro His His Pro His Ala His Pro
225                 230                 235                 240

His Asn Pro Gln Pro Met His Arg Tyr Asp Met Gly Ala Leu Gln Tyr
                245                 250                 255

Ser Pro Ile Ser Asn Ser Gln Gly Tyr Met Ser Ala Ser Pro Ser Gly
            260                 265                 270

Tyr Gly Gly Leu Pro Tyr Gly Ala Ala Ala Ala Ala Ala Ala Ala His
        275                 280                 285

Gln Asn Ser Ala Val Ala Ala Ala Ala Ala Ala Ala Ala Ala Ser Ser
    290                 295                 300

Gly Ala Leu Gly Ala Leu Gly Ser Leu Val Lys Ser Glu Pro Ser Gly
305                 310                 315                 320

Ser Pro Pro Ala Pro Ala His Ser Arg Ala Pro Cys Pro Gly Asp Leu
                325                 330                 335

Arg Glu Met Ile Ser Met Tyr Leu Pro Ala Gly Glu Gly Gly Asp Pro
            340                 345                 350

Ala Ala Ala Ala Ala Ala Ala Ala Gln Ser Arg Leu His Ser Leu Pro
        355                 360                 365

Gln His Tyr Gln Gly Ala Gly Ala Gly Val Asn Gly Thr Val Pro Leu
    370                 375                 380

Thr His Ile  *
385
```

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1161 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: RNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(7..9, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(28..30, "cun")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(34..36, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(64..66, "cun")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(67..69, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(148..150, "cgn")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(157..159, "cgn")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(184..186, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(187..189, "cgn")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(196..198, "cgn")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(199..201, "cgn")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(235..237, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(250..252, "cgn")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(244..246, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(253..255, "cun")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(277..279, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(291..294, "cgn")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(316..318, "cgn")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(319..321, "cun")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(322..324, "cgn")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(328..330, "cun")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(361..363, "cgn")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(367..369, "cgn")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(370..372, "cgn")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(385..387, "cun")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(388..390, "cun")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(406..408, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(409..411, "cun")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(421..423, "cun")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(424..426, "cun")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(505..507, "cgn")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(508..510, "cun")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(513..515, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(583..585, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(631..632, "cun")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(676..678, "cgn")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(742..744, "cgn")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(760..762, "cun")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(769..771, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(778..780, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(784..786, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(799..801, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(805..807, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(811..813, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(836..838, "cun")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(871..873, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(907..909, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(910..912, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(919..921, "cun")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(928..930, "cun")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(934..936, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(937..939, "cun")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(946..948, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(955..957, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(961..963, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(982..984, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(985..987, "cgn")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(1006..1008, "cun")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(1009..1011, "cgn")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(1021..1023, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(1030..1032, "cun")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(1084..1086, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(1087..1089, "cgn")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(1090..1092, "cun")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(1096..1098, "agy")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(1099..1101, "cun")

(ix) FEATURE:
(A) NAME/KEY: variation
(B) LOCATION:replace(1150..1152, "cun")

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

AUGUAYUCNA UGAUGAUGGA RACNGAYUUR CAYUCNCCNG GNGGNGCNCA RGCNCCNACN 60

AAYUURUCNG GNCCNGCNGG NGCNGGNGGN GGNGGNGGNG GNGGNGGNGG NGGNGGNGGN 120

GGNGGNGGNG CNAARGCNAA YCARGAYAGR GUNAARAGRC CNAUGAAYGC NUUYAUGGUN 180

UCGUCNAGRG GNCARAGRAG RAARAUGGCN CARGARAAYC CNAARAUGCA YAAYUCNGAR 240

AUHUCNAARA GRUURGGNGC NGARUCGAAR GUNAUGUCNG ARGCNGARAA RAGRCCNUUY 300

AUHGAYGARG CNAARAGRUU RAGRGCNUUR CAYAUGAARG ARCAYCCNGA YUAYAARUAY 360

AGRCCNAGRA GRAARACNAA RACNUURUUR AARAARGAYA ARUAYUCNUU RGCNGGNGGN 420

UURUURGCNG CNGGNGCNGG NGGNGGNGGN GCNGCNGUNG CNAUGGGNGU NGGNGUNGGN 480

GUNGGNGCNG CNCCNGUNGG NCARAGRUUR GARUCNCCNG GNGGNGCNGC NGGNGGNGCN 540

UAYGCNCAYG UNAAYGGNUC GGCNAAYGGN GCNUAYCCNG GNUCNGUNGC NGCNGCNGCN 600

GCNGCNGCNG CNAUGAUGCA RGARGCNCAR UURGCNUAYG GNCARCAYCC NGGNGCNGGN 660

GGNGCNCAYC CNCAYAGRAC NCCNGCNCAY CCNCAYCCNC AYCAYCCNCA YGCNCAYCCN 720

CAYAAYCCNC ARCCNAUGCA YAGRUAYGAY AUGGGNGCNU URCARUAYUC NCCNAUHUCN 780

AAYUCNCARG GNUAYAUGUC NGCNUCNCCN UCNGGNUAYG GNGGNUURCC NUAYGGNGCN 840

GCNGCNGCNG CNGCNGCNGC NCAYCARAAY UCNGCNGUNG CNGCNGCNGC NGCNGCNGCN 900

GCNGCNUCNU CNGGNGCNUU RGGNGCNUUR GGNUCNUURG UNAARUCNGA RCCNUCNGGN 960

UCNCCNCCNG CNCCNGCNCA YUCNAGRGCN CCNUGYCCNG GNGAYUURAG RGARAUGAUH 1020

UCNAUGUAYU URCCNGCNGG NGARGGNGGN GAYCCNGCNG CNGCNGCNGC NGCNGCNGCN 1080

CARUCNAGRU URCAYUCNUU RCCNCARCAY UAYCARGGNG CNGGNGCNGG NGUNAAYGGN 1140

ACNGUNCCNU URACNCAYAU H 1161

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

7

WWCAAWG

Claims

1. A method for isolating one or more neuroblastic cells from a population of cells comprising the steps of:

(a) detecting the expression of the *Sox1* gene in the cell(s); and

(b) sorting the cell(s) to isolate those cells expressing the *Sox1* gene.

2. A method according to claim 1, wherein the population of cells is derived from CNS tissue.

3. A method according to claim 1, wherein the population of cells is derived from a cell culture.

4. A method according to any preceding claim, wherein the expression of the *Sox1* gene is detected by nucleic acid hybridisation.

5. A method according to any one of claims 1 up to 3, wherein the expression of the *Sox1* gene is detected by a binding of SOX1 to a detectable ligand.

6. A method according to claim 5, wherein the detectable ligand is a labelled immunoglobulin.

7. A method according to claim 5, wherein the detectable ligand is a labelled oligonucleotide complementary to *Sox1* mRNA.

8. A method according to any preceding claim, wherein the expression of the *Sox1* gene is detected by FACS analysis.

9. A method for isolating a neuroblastic cell from a population of cells, comprising the steps of:

(a) transfecting the population of cells with a genetic construct comprising a coding sequence encoding a detectable marker operatively linked to a *Sox1* control region;

(b) detecting the cells which express the selectable marker; and

(c) sorting the cells which express the selectable marker from the population of cells.

10. A method for isolating a neuroblastic cell from a population of cells, comprising the steps of:

(a) transfecting the population of cells with a genetic construct comprising a coding sequence encoding a detectable marker operatively linked to a control sequence which is transactivatable by SOX1;

(b) detecting the cells which express the selectable marker; and

(c) sorting the cells which express the selectable marker from the population of cells.

11. A method according to claim 9 or claim 10, wherein the selectable marker is a fluorescent or luminescent polypeptide.

12. A method according to claim 9 or claim 10, wherein the selectable marker is a polypeptide detectable at the surface of the cell.

13. A method according to claim 9, wherein the *Sox1* control sequence comprises nucleotides 1 to 60 of SEQ ID No. 3.

14. A method according to claim 10, wherein the element transactivatable by SOX1 comprises the sequence motif A/T A/T CAA A/T G.

15. A method for producing a cell committed to the neuronal lineage, comprising the steps of:

(a) transfecting a pluripotent stem cell with a genetic construct comprising a coding sequence expressing *Sox1*;

(b) culturing the stem cells in order to differentiate them into neural cells; and

(c) isolating the neural cells thereby produced.

16. A method according to claim 15, wherein the *Sox1* sequence is operatively linked to an inducible promoter.

17. A method according to claim 15 or claim 16, wherein the cell is further transfected with a vector comprising a sequence encoding a regulator which modulates the expression of the *Sox1* sequence.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : **C12N 5/06, 5/08, C12Q 1/68, C12N 5/10, 15/85**

A3

(11) International Publication Number: **WO 99/00516**

(43) International Publication Date: 7 January 1999 (07.01.99)

(21) International Application Number: PCT/GB98/01862

(22) International Filing Date: 25 June 1998 (25.06.98)

(30) Priority Data:
9713469.6 25 June 1997 (25.06.97) GB

(71) Applicant *(for all designated States except US)*: MEDICAL RESEARCH COUNCIL [GB/GB]; 20 Park Crescent, London W1N 4AL (GB).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: LOVELL-BADGE, Robin [GB/GB]; National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA (GB). PEVNY, Laryssa, H. [GB/GB]; National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA (GB). SMITH, Austin [GB/GB]; Centre for Genome Research, University of Edinburgh, East Mains Road, Edinburgh EH9 3WQ (GB).

(74) Agents: MASCHIO, Antonio et al.; D. Young & Co., 21 New Fetter Lane, London EC4A 1DA (GB).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(88) Date of publication of the international search report:
18 March 1999 (18.03.99)

(54) Title: NEURONAL STEM CELL GENE

(57) Abstract

The invention relates to a method for isolating a neuroblastic cell from a population of cells comprising the steps of: (a) detecting the expression of the *Sox1* gene in the cells; and (b) sorting the cells to isolate those cells expressint the *Sox1* gene, as well as to a method for causing a cell to become committed to the neural pathway comprising administering SOX1 thereto.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

INTERNATIONAL SEARCH REPORT

Intern ıal Application No
PCT/GB 98/01862

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N5/06 C12N5/08 C12Q1/68 C12N5/10 C12N15/85

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 97 02049 A (EMORY UNIVERSITY) 23 January 1997 see the whole document ---	1-17
A	US 5 639 618 A (GAY D.A.) 17 June 1997 see the whole document ---	1-17
A	PEVNY L H ET AL: "Sox genes find their feet." CURRENT OPINION IN GENETICS AND DEVELOPMENT, (1997 JUN) 7 (3) 338-44, XP002087357 see the whole document --- -/--	1-17

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
10 December 1998	28/12/1998

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Moreau, J

1

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Interr ıal Application No
PCT/GB 98/01862

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category ˇ	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
P,X	REX M ET AL: "cSox21 exhibits a complex and dynamic pattern of transcription during embryonic development of the chick central nervous system." MECHANISMS OF DEVELOPMENT, (1997 AUG) 66 (1-2) 39-53, XP002087358 see the whole document ---	1-17
P,X	DATABASE MEDLINE US NATIONAL LIBRARY OF MEDICINE (NLM), BETHESDA, MD, US PEVNY L H ET AL: "A role for SOX1 in neural determination." XP002087359 see abstract & DEVELOPMENT, (1998 MAY) 125 (10) 1967-78, ENGLAND: United Kingdom -----	1-17

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

...ormation on patent family members

Interr ...al Application No
PCT/GB 98/01862

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9702049 A	23-01-1997	US 5753505 A	19-05-1998
		AU 6452196 A	05-02-1997
		CA 2226417 A	23-01-1997
		EP 0841950 A	20-05-1998
US 5639618 A	17-06-1997	NONE	